The Bell System Technical Journal
Vol. 54, No. 9, November 1975
Printed in U.S.A.

A Diffusion Model Approximation for the GI/G/1 Queue in Heavy Traffic

By D. P. HEYMAN

(Manuscript received April 10, 1975)

For the single-server queue with renewal input, we obtain heavy traffic approximations for the time-dependent distributions of queue length and virtual delay by constructing approximating diffusion processes. These approximations are shown to agree with known limiting cases, and a comparison is made with results from a computer simulation.

I. INTRODUCTION

We consider a single-server queuing system where the interarrival times are independent and identically distributed (i.i.d.) random variables, customers are served in order of arrival, the service times of the various customers are i.i.d. random variables, and the interarrival and service times form independent sequences. Lindley[1] obtained a recursive equation for the delay-time of the nth arriving customer, an integral equation for the delay time of a customer in the steady-state, and conditions for the latter to have a nondegenerate limit.

Lindley's equations have not yielded to conveniently used analytical solutions, except in some special cases, stimulating a search for approximations to the distributions of general interest. In this paper, we approximate the queue length and delay processes by appropriately chosen diffusion processes. This method of approximation appears to have been introduced by Gaver[2] and Newell.[3] Gaver and Newell considered the M/G/1 queue; we extend their approximate models to the GI/G/1 queue. Other methods for obtaining diffusion approximations for queuing processes involve applying the theory of weak convergence to sequences of approximating processes. Whitt[4] is a survey of these methods and contains an extensive bibliography. We show that the diffusion models developed in this paper agree with the limiting processes obtained by weak convergence methods.

Important features of this paper are the use of the M/M/1 queue to motivate a diffusion process approximation for the single-server queue and the use of elementary renewal theory results to obtain the param-

eters of the process. This approach provides an intuitive explanation for the limit theorems.

II. PRELIMINARIES AND NOTATION

Let T_i be time between the arrival epochs of the $(i-1)$st and ith customers, $i = 1, 2, \cdots$, assume these random variables are i.i.d., and let $\lambda^{-1} = E(T_1)$ and $\sigma_A^2 = \text{Var}\,(T_1)$. Assume the customers are served in order of arrival, let S_i be the service time of the ith customer, $\mu^{-1} = E(S_1)$, $\sigma_B^2 = \text{Var}\,(S_1)$, and assume $S_1, S_2, \cdots$ are i.i.d. random variables. We define the traffic intensity by $\rho = \lambda/\mu$ and will always assume $\rho < 1$. We seek approximations for the queue size and virtual delay at time t, and obtain these approximations from suitably chosen diffusion processes. For any function F, let $F_x = \partial F/\partial x$, $F_{xx} = \partial^2 F/\partial x^2$, $F_y = \partial F/\partial y$, etc., and unambiguous arguments will be suppressed.

Let $\{X(t), t \geqq 0\}$ be a homogeneous and additive diffusion process, $F(t, x; x_0) = \Pr\{X(t) \leqq x \mid X(0) = x_0\}$, and a and b be the infinitesimal mean and variance of the process, respectively. Then F satisfies the forward Kolmogorov (Fokker-Planck) equation

$$F_t = -aF_x + \frac{b}{2} F_{xx}, \tag{1}$$

with initial condition

$$F(0, x; x_0) = \begin{Bmatrix} 0 & \text{if} & x < x_0 \\ 1 & \text{if} & x \geqq x_0 \end{Bmatrix}. \tag{2}$$

If the range of $X(t)$ is $[0, \infty)$, then (1) is subject to the boundary condition

$$F(t, 0; x_0) = 0, \qquad t > 0. \tag{3}$$

The solution to (1) subject to (2) and (3) is given in Newell[5] as

$$F(x, t; x_0) = \Phi\left(\frac{x - x_0 - at}{\sqrt{bt}}\right) - e^{2xa/b}\Phi\left(\frac{-x - x_0 - at}{\sqrt{bt}}\right), \tag{4}$$

where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du$$

is the normal distribution function (d.f.). If $a < 0$, then

$$F(x) = \lim_{t \to \infty} F(x, t; x_0) = 1 - e^{-2x(-a)/b}, \tag{5}$$

which is the negative-exponential d.f. and is independent of x_0.

Returning to the GI/G/1 queue, let W_i denote the delay of the ith customer. Lindley showed that when $\rho < 1$, $W = \lim_{i \to \infty} W_i$ is nondegenerate. Kingman[6] showed that if ρ is slightly less than unity, the

d.f. of W is approximately negative-exponential with mean

$$\tfrac{1}{2}(\sigma_A^2 + \sigma_B^2)/(\lambda^{-1} - \mu^{-1}). \tag{6}$$

III. AN APPROXIMATION FOR THE M/M/1 QUEUE

To motivate the diffusion model employed in approximating the GI/G/1 queue, and to indicate its efficacy for the M/M/1 queue in particular, we first develop an approximation for the M/M/1 queue. A scbeme suggested by Bailey[7] for approximating stocbastic processes is used.

In tbe M/M/1 queue, customers arrive according to a Poisson process with rate λ, and the service-time distribution is negative-exponential with mean μ^{-1}. Let $N(t)$ denote the number of customers in the queue at time t, and $\pi(t, n; n_0) = \Pr\{N(t) = n | N(0) = n_0\}$. For $t > 0$ and $n = 1, 2, \cdots$, $\pi(t, n; n_0)$ satisfies

$$\begin{aligned}\frac{d}{dt}\pi(t, n; n_0) = \lambda\pi(t, n-1; n_0) &+ \mu\pi(t, n+1; n_0) \\ &- (\lambda + \mu)\pi(t, n; n_0), \end{aligned} \tag{7a}$$

and

$$\frac{d}{dt}\pi(t, 0; n_0) = \mu\pi(t, 1; n_0) - \lambda\pi(t, 0; n_0). \tag{7b}$$

The initial condition is

$$\pi(0, n; n_0) \begin{Bmatrix} 1 & \text{if} & n = n_0 \\ 0 & \text{if} & n \neq n_0 \end{Bmatrix}, \tag{8}$$

and the boundary condition is

$$\pi(t, n; n_0) = 0 \quad n < 0, \qquad t \geqq 0. \tag{9}$$

The idea of the approximation is to replace (7) hy a partial differential equation that is easier to solve. We do this hy replacing tbe discrete variable n by the continuous variable x, and $\pi(t, n; n_0)$ hy $p(t, x; x_0)$ in (7). Expanding in a Taylor's series about the point $(t, x; x_0)$ and keeping only first- and second-order terms, we obtain

$$p_t = -(\lambda - \mu)p_x + \frac{\lambda + \mu}{2} p_{xx}, \qquad x, t > 0. \tag{10}$$

If we define

$$\mathrm{P}(t, x; x_0) = \int_{-\infty}^{x} p(t, y; x_0)dy,$$

it can he easily shown that P also satisfies (10). We take

$$\mathrm{P}(0, x; x_0) = \begin{Bmatrix} 0 & \text{if} & x < x_0 \\ 1 & \text{if} & x \geqq x_0 \end{Bmatrix} \tag{11}$$

and

$$P(t, 0; x_0) = 0, \qquad t > 0, \tag{12}$$

as natural replacements for (8) and (9). This system of equations is identical in form to (1), (2), and (3), so $P(t, x; x_0)$ is given by the right-side of (4) with $a = \lambda - \mu$ and $b = \lambda + \mu$.

Consider now the asymptotic behavior of π and P: From the theory of the M/M/1 queue, we have

$$\pi_n = \lim_{t \to \infty} \pi(t, n; n_0) = (1 - \rho)\rho^{n-1} \quad \text{for} \quad n > 0$$

and from (5) we obtain

$$P(t) = \lim_{t \to \infty} P(t, x; x_0) = 1 - \exp\left[-2(1 - \rho)/(1 + \rho)\right].$$

If for $n > 0$ we approximate π_n by

$$\tilde{\pi}_n = \int_{n-1}^{n} dP(t),$$

we obtain

$$\begin{aligned} \tilde{\pi}_n &= \left[1 - e^{-2(1-\rho)/(1+\rho)}\right] e^{-2(n-1)(1-\rho)/(1+\rho)} \\ &= (1 - \alpha)\alpha^{n-1}, \qquad n > 0, \end{aligned}$$

where $\alpha = \exp\left[-2(1 - \rho)/(1 + \rho)\right]$, so $\tilde{\pi}_n$ has the same form as π_n. If ρ is close to unity, $\alpha \doteq \rho$ and hence

$$\tilde{\pi}_n \doteq \pi_n, \qquad \rho \doteq 1,$$

so $\tilde{\pi}_n$ is a good approximation of π_n when ρ is slightly less than one.

IV. THE GI/G/1 QUEUE—APPROXIMATE QUEUE LENGTH

Let us first consider a heuristic manner of obtaining (10) for the M/M/1 queue. During the time interval $(t, t + \Delta t]$, the number of customers in the system changes by the number of arrivals minus the number of service completions, and when $N(t) = n > 0$, this change has expectation $(\lambda - \mu)\Delta t + o(\Delta t)$ and variance $(\lambda + \mu)\Delta t + o(\Delta t)$. To approximate $\{N(t), t \geqq 0\}$ by a diffusion process with the same infinitesimal mean and variance, set $a = \lambda - \mu$ and $b = \lambda + \mu$ in (1), which yields (10). This suggests that an appropriate choice of a and b will yield a good approximation for the queue length of the GI/G/1 queue.

For the GI/G/1 queue, let $N(t)$ be the queue size at time t, $A(t)$ and $D(t)$ the number of arrivals and departures in $(0, t]$, respectively; then

$$N(t) = N(0) + A(t) - D(t), \qquad t > 0. \tag{13}$$

For any renewal process $\{M(t), t \geqq 0\}$ where the interevent times have mean m and variance V, for large values of t

$$E[M(t)] \approx t/m, \tag{14}$$

and

$$\mathrm{Var}\,[M(t)] \approx tV/m^3, \tag{15}$$

(Ref. 8). By hypothesis $\{A(t), t \geqq 0\}$ is a renewal process, so from (14) and (15) we obtain

$$E[A(t)] \approx \lambda t, \qquad \mathrm{Var}\,[A(t)] \approx \lambda^3 \sigma_A^2 t. \tag{16}$$

The process $\{D(t), t \geqq 0\}$ is not a renewal process, but, in heavy traffic (ρ close to 1), the server will be occupied most of the time, so we approximate $D(t)$ by $\tilde{D}(t)$, where

$$E[\tilde{D}(t)] \approx \mu t, \qquad \mathrm{Var}\,[\tilde{D}(t)] \approx \mu^3 \sigma_B^2 t. \tag{17}$$

Cox and Smith[9] use (13), (16), and (17) to study the GI/G/1 queue for small values of t without using a diffusion model. Substituting $\tilde{D}(t)$ for $D(t)$ in (13) and using (16) and (17), we obtain the approximate results

$$\lim_{t \to \infty} E[N(t)]/t \doteq \lambda - \mu \tag{18}$$

and

$$\lim_{t \to \infty} \mathrm{Var}\,[N(t)]/t \doteq \lambda^3 \sigma_A^2 + \mu^3 \sigma_B^2, \tag{19}$$

which suggest that we approximate $N(t)$ by a diffusion process $\tilde{N}(t)$, say, with infinitesimal mean and variance given by

$$a = \lambda - \mu \tag{20}$$

and

$$b = \lambda^3 \sigma_A^2 + \mu^3 \sigma_B^2, \tag{21}$$

respectively. If we let $F(t, x; x_0) = \Pr\{\tilde{N}(t) \leqq x \mid \tilde{N}(0) = x_0\}$, then F satisfies (1), (2), (3), and hence is given by (4), with a and b as above.

As a partial check on the efficacy of this approximation, let us define $\tilde{N} = \lim_{t \to \infty} \tilde{N}(t)$. For $\rho < 1$, $\tilde{N}$ is a proper random variable, and from (5), (20), and (21),

$$E(\tilde{N}) = \frac{\mu}{2} \frac{\sigma_A^2 \rho^2 + \sigma_B^2 \rho^{-1}}{\lambda^{-1} - \mu^{-1}}, \tag{22}$$

which together with the queuing formula $L = \lambda W$ (see Ref. 10) and $\rho = 1$ yields the heavy traffic approximation given by (6).

V. THE GI/G/1 QUEUE—APPROXIMATE VIRTUAL DELAY

The virtual delay at time t is the delay in queue a customer would experience if it arrived at time t; an exact formulation is given by

Beneš.[11] Toward developing an approximation of the virtual-delay process, define

$$L(t) = S_1 + S_2 + \cdots + S_{A(t)}, \qquad t > 0, \tag{23}$$

so $L(t)$ represents the amount of work time hrought to the server in $(0, t]$. Since $S_1, S_2, \cdots, S_{A(t)}$ are i.i.d., using (17) we obtain

$$E[L(t)] = E(S_1)E[A(t)] \approx \rho t \tag{24}$$

for large values of t. Using the conditional variance relationship $\text{Var}\,[L(t)] = E\{\text{Var}\,[L(t)\,|\,A(t)]\} + \text{Var}\,\{E[L(t)\,|\,A(t)]\}$ and (17), we ohtain for large values of t,

$$\text{Var}\,[L(t)] \approx \lambda(\sigma_B^2 + \rho^2\sigma_A^2)t. \tag{25}$$

The sample paths of the virtual delay process are sawtooth functions with a jump of size S_i at the arrival epoch of the ith customer followed by a decline of slope -1; the process has an impenetrahle houndary at the axis of ahscissas. Assume that (24) and (25) hold for all t, so that

$$\alpha = \lim_{\Delta t \to 0} \{E[L(t + \Delta t)] - E[L(t)]\}/\Delta t = \rho \tag{26}$$

and

$$\begin{aligned} \beta &= \lim_{\Delta t \to 0} \{\text{Var}\,[L(t + \Delta t)] - \text{Var}\,[L(t)]\}/\Delta t \\ &= \lambda(\sigma_B^2 + \rho^2\sigma_A^2). \end{aligned} \tag{27}$$

Following Gaver,[2] we approximate the virtual-delay process hy a diffusion process $\{\tilde{V}(t), t \geqq 0\}$, say, with infinitesimal mean and variance given hy

$$a = \rho - 1, \qquad b = \lambda(\sigma_B^2 + \rho^2\sigma_A^2), \tag{28}$$

respectively. Hence, the time-dependent d.f. of $\tilde{V}(t)$ is given by (4) with a and b given by (28).

Turning now to asymptotic results, when $\rho < 1$, $\tilde{V} = \lim_{t\to\infty} \tilde{V}(t)$ exits and is proper, and from (5) has a negative-exponential distrihution with mean

$$E(\tilde{V}) = \tfrac{1}{2}(\sigma_B^2 + \rho^2\sigma_A^2)/(\lambda^{-1} - \mu^{-1}). \tag{29}$$

Hooke[12] showed that, if the interarrival times are nonlattice, then as $\rho \uparrow 1$, $\Pr\,\{W \leqq x\} \to \Pr\,\{V \leqq x\}$, where V is the virtual delay in the steady state. This result and (29) together show that, as $\rho \uparrow 1$, $E(\tilde{V})/E(W) \uparrow 1$, and the d.f. of $\tilde{V}$ agrees with (6). From (22) and (29) we ohserve that $E(\tilde{V}) < E(\tilde{N})/\mu$ when $\rho < 1$, but $E(V) \geqq E(N)/\mu$ is known to hold.

We note that once a diffusion model for the virtual-delay process is at hand, the method given in Heyman[13] for approximating the busy-period distribution can he applied.

VI. A NUMERICAL EXAMPLE

In this section, we compare $E[N(t)]$ to the results of a single computer simulation. We consider a single-server queue that is empty at time zero, where the interarrival times are uniformly distrihuted from 0 to 20 minutes, and the service times are uniformly distrihuted from 0 to 19 minutes. Thus,

$$\lambda^{-1} = 10, \qquad \sigma_A^2 = 100/3, \qquad \mu^{-1} = 9.5, \qquad \sigma_B^2 = (9.5)^2/3; \qquad (30)$$

hence, $\rho = 0.95$ and from (22) we ohtain $E(\tilde{N}) = 6.50$. Since the d.f. of $\tilde{N}$ is negative-exponential, the standard deviation of $\tilde{N}$ is also 6.50.

Let $N_j(t)$ denote the sample of $N(t)$ produced hy the jth simulation run, N_j he the estimate of N produced hy the jth run, and $\hat{N} = n^{-1} \sum_{j=1}^{n} N_j$. Assuming the simulator produces independent sample paths and approximating the mean and variance of N hy those of $\tilde{N}$, standard sampling theory indicates that, for reasonahly large values of n, $\hat{N}$ has a normal distribution with mean 6.5 and standard deviation $6.5/\sqrt{n}$. Hence, given $c > 0$, the numher of runs required to have

$$\Pr \{ |\hat{N} - E(N)| \leqq cE(N) \} \geqq 0.99$$

is the smallest integer $\geqq (2.575)^2/c$; choosing $c = 0.05$ yields $n = 134$.

These considerations lead us to simulate 150 sample paths, and indicate one of the potential uses of even a crude diffusion approximation.

Gaver[2] and Newell[5] observe that if one makes the change of variables $\xi = -(a/b)x$ and $\tau = (a^2/b)t$, (1) hecomes

$$F_\tau = -F_\xi + F_{\xi\xi}/2, \qquad (31)$$

which can he solved once and for all; the solutions for any a and b can he recovered hy scaling. For our example, $-(b/a) = 13.00$ and $b/a^2 = 2472.2$ minutes. Table 2-2 in Gaver[2] gives the values of

$$\int_0^\infty \xi d_\xi F(\tau, \xi; 0)$$

for $\tau = 0.1, 0.2, \cdots, 1.0$. These were used to construct Tahle I helow.

The approximations shown in the table are not as accurate as the diffusion approximations for M/G/1 shown in Tahle I of Gaver[2]; these comparisons share with Gaver's table the property that the diffusion process consistently overestimates the mean queue-size.

Table I — Comparison of diffusion approximation and simulation results

Time	τ									
$E[\tilde{N}(t)]$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Diffusion	2.68	3.50	4.02	4.40	4.68	4.91	5.11	5.27	5.41	5.52
Simulation	2.25	2.91	3.40	3.71	4.16	4.39	4.61	5.05	5.18	5.13

VII. COMPARISON WITH THE RESULTS OF IGLEHART AND WHITT

Iglehart and Whitt[14,15] use the theory of weak convergence of stochastic processes to ohtain heavy-traffic-limit theorems for large classes of queuing processes. In particular, these theorems hold for the GI/G/1 queue. Their results are obtained by considering a sequence of GI/G/1 queuing systems, indexed hy n ($n = 1, 2, \cdots$). For the nth system, let λ_n and μ_n denote the arrival and service rates, respectively, and $N_n(t)$ and $V_n(t)$ denote the queue size and virtual delay at time t, respectively. For the ahove quantities, let the ahsence of a suhscript denote the limit with respect to n, e.g., $\lambda = \lim_{n\to\infty} \lambda_n$. For each n, let σ_A^2 and σ_B^2 be as hefore, and assume $N_n(0) = V_n(0) = 0$.

Section three of Iglehart and Whitt[14] shows that if $\lim_{n\to\infty} (\lambda_n - \mu_n)\sqrt{n} = c$, where c is some finite constant, then

$$\lim_{n\to\infty} \frac{N_n(nt')}{\gamma\sqrt{n}} \Rightarrow B(t', -c/\gamma), \tag{32}$$

where $\gamma = \lambda^3\sigma_A^2 + \mu^3\sigma_B^2$, $B(t', -c/\gamma)$ is the Wiener process with negative drift c/γ together with an impenetrahle harrier at the origin and $\Rightarrow$ denotes weak convergence. From (32) it follows that for large values of n, i.e., when $\rho_n = \lambda_n/\mu_n$ is close to 1,

$$\Pr\{N_n(nt')/\gamma\sqrt{n} \leqq d\} = \Phi\left(\frac{d - ct'/\gamma}{\sqrt{t'}}\right) - e^{2cd/\gamma}\Phi\left(\frac{-d - ct'/\gamma}{\sqrt{t'}}\right). \tag{33}$$

Since $N_n(t) \doteq N(t)$, $c \doteq (\lambda - \mu)\sqrt{n}$ for large n, upon making the change of variahles $x = d\gamma\sqrt{n}$ and $t = nt'$, we ohtain

$$\Pr\{N(t) \leqq x\} \doteq \Phi\left(\frac{x - (\lambda - \mu)t}{\gamma\sqrt{t}}\right) - e^{2(\lambda-\mu)x/\gamma^2}\Phi\left(\frac{-x - (\lambda - \mu)t}{\gamma\sqrt{t}}\right) \tag{34}$$

from (33). Observe that the right-hand side of (34) agrees exactly with

the time-dependent distribution (when $x_0 = 0$) of $\tilde{N}(t)$ obtained in Section IV.

Let $\alpha_n = \mu_n^{-1} - \lambda_n^{-1}$ and $\sigma^2 = \sigma_A^2 + \sigma_B^2$. From theorem 6.1 of Iglebart and Whitt,[15] it is possible to derive tbe following result, which appears in theorem 4 of Whitt.[16] If $\alpha_n \sigma^{-1}\sqrt{n} \to -k$, $-\infty < k < \infty$, then

$$\frac{V_n(nt')}{\sigma\sqrt{n\lambda_n}} \Rightarrow B(t', -k). \tag{35}$$

For large values of n, $\alpha_n \doteq (\rho - 1)\lambda^{-1}$, $-k \doteq (\rho - 1)\lambda^{-1}\sigma^{-1}\sqrt{n}$, $V_n(t) \doteq V(t)$, and

$$\Pr\{V(nt') \leqq d\sigma\sqrt{\lambda n}\} \doteq \Phi\left(\frac{d + kt'\sqrt{\lambda}}{\sqrt{t'}}\right) - e^{-2kd\sqrt{\lambda}}\Phi\left(\frac{-d + kt'\sqrt{\lambda}}{\sqrt{t'}}\right). \tag{36}$$

Letting $t = nt'$ and $x = d\sigma\sqrt{\lambda n}$ in (35), we obtain

$$\Pr\{V(t) \leqq x\} \doteq \Phi\left(\frac{x - (\rho - 1)t}{\sigma\sqrt{\lambda t}}\right) - e^{2(\rho-1)x/\lambda\sigma^2}\Phi\left(\frac{-x - (\rho - 1)t}{\sigma\sqrt{\lambda t}}\right). \tag{37}$$

Tbe right-hand side of (36) represents tbe time-dependent distribution (4) with $x_0 = 0$ and

$$a = \rho - 1, \qquad b = \lambda(\sigma_A^2 + \sigma_B^2), \tag{38}$$

which is the same value of a and almost the same value of b given in (28), where the difference between tbe values of b given in (28) and (38) vanisbes as $\rho \to 1$.

From tbe results of this section, we can conclude tbat the heuristically constructed diffusion models for $\tilde{N}(t)$ and $\tilde{V}(t)$ given in this paper can be regarded as the limit (in the sense of weak convergence) of $N(t)$ and $V(t)$, respectively, with suitable normalization.

VIII. SUMMARY

We have obtained an approximation for the time-dependent distributions of queue lengtb and virtual delay in a GI/G/1 queue using a diffusion model. The diffusion model for the queue length process was obtained by using the mean and variance of the asymptotic rate of change of an approximation of the queue-length process as the infinitesimal mean and variance for a diffusion process. The diffusion approximation was compared to a simulation of a particular queueing process, and reasonably close agreement for mean queue lengths was obtained. The approximation for the virtual-delay process was gen-

erated in a manner suggested in Gaver[2] and the limiting results it provided were shown to agree with a theorem of Kingman[6] for the delay process. The time-dependent distributions for the approximate queue length and virtual-delay processes agree, as the traffic intensity approaches one, with the limiting results of Iglehart and Whitt.[14,15]

IX. ACKNOWLEDGMENT

John Rath of Bell Laboratories suggested Section VII and made many helpful comments.

REFERENCES

1. D. V. Lindley, "The Theory of Queues with a Single Server," Proc. Comb. Phil. Soc., *48*, Part 2 (April 1952), pp. 277–289.
2. D. P. Gaver, "Diffusion Approximations and Models for Certain Congestion Problems," J. Appl. Prob., *5*, No. 3 (December 1968), pp. 607–623.
3. G. F. Newell, "Queues with Time-Dependent Arrival Rates I—The Transition Through Saturation," J. Appl. Prob., *5*, No. 2 (August 1968), pp. 436–451.
4. W. Whitt, "Heavy Traffic Limit Theorems for Queues: A Survey," *Mathematical Methods in Queueing Theory*, New York: Springer-Verlag, 1974, pp. 307–350.
5. G. F. Newell, *Applications of Queueing Theory*, London: Chapman and Hall, 1971.
6. J. F. C. Kingman, "The Heavy Traffic Approximation in the Theory of Queues," *Proceedings of the Symposium on Congestion Theory*, W. L. Smith and W. E. Wilkinson eds., Chapel Hill: University of North Carolina Press, 1965, pp. 137–159.
7. N. T. J. Bailey, *The Elements of Stochastic Processes*, New York: John Wiley, 1964.
8. D. R. Cox, *Renewal Theory*. London: Methuen, 1962.
9. D. R. Cox and W. L. Smith, *Queues*. London: Methuen, 1961.
10. J. D. C. Little, "A Proof of the Queueing Formula $L = \lambda W$," Opns. Res., *9*, No. 3 (May–June 1961), pp. 383–387.
11. V. E. Beneš, *General Stochastic Processes in the Theory of Queues*, Reading, Massachusetts: Addison–Wesley, 1963.
12. J. A. Hooke, "Some Limit Theorems for Priority Queues," Tech. Report 91, Ithaca: Cornell University, 1969.
13. D. P. Heyman, "An Approximation for the Busy-Period of the M/G/1 Queue Using a Diffusion Model," J. Appl. Prob., *11*, No. 1 (March 1974), pp. 159–169.
14. D. Iglehart and W. Whitt, "Multiple Channel Queues in Heavy Traffic, I," Adv. Appl. Prob., *2*, No. 1 (Spring 1970), pp. 150–177.
15. D. Iglehart and W. Whitt," Multiple Channel Queues in Heavy Traffic, II: Sequences, Networks, and Batches," Adv. Appl. Prob., *2*, No. 2 (Autumn 1970), pp. 355–369.
16. W. Whitt, "Heavy Traffic Approximations for Stable Queues," Technical Report, Department of Administrative Sciences, New Haven: Yale University, August 1971.